MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

A. I. Memo 322 March 1975

A Frame for Frames: Representing Knowledge for Recognition*

Benjamin J. Kuipers

ABSTRACT

This paper presents a version of frames suitable for representing knowledge for a class of recognition problems. An initial section gives an intuitive model of frames, and illustrates a number of desirable features of such a representation. A more technical example describes a small recognition program for the Blocks World which implements some of these features. The final section discusses the more general significance of the representation and the recognition process used in the example.

* This is a preprint of a paper to appear in Daniel G. Bobrow & Allan M. Collins (Eds.), Representation and Understanding, New York: Academic Press, 1975.

This report describes research done in part at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-70-A-0362-0003.

A FRAME FOR FRAMES: REPRESENTING KNOWLEDGE FOR RECOGNITION

Benjamin J. Kuipers
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

I. INTRODUCTION

How can we represent in a computer program the kind of knowledge people manipulate easily and effectively? One of the significant discoveries of artificial intelligence has been how computationally difficult are the elementary tasks of vision, language, and common sense reasoning which we perform continually in the course of our everyday activities. The techniques used by the artificially intelligent programs of the past decade are simply not powerful enough to approach human performance over any wide range of tasks. New mechanisms have recently been proposed by which the organization of previously accumulated knowledge can assist active perception and understanding. Briefly, the idea is that if there is too little computation time when a problem comes up, do some of the work in advance and keep the computed results available. This in itself is not an astonishing insight, though it does focus our attention on the relationship among immediate perception, understanding, and long-term, real-world knowledge. It obviously should be easier to see something which has previously been seen, and the question becomes how to organize and use such previous experience.

Minsky (1975) proposed a theory of "frames" as a mechanism for representing knowledge in the computer. A frame is a structure which represents knowledge about a very limited domain. A frame produces a description of the object or action in question, starting with an invariant structure common to all cases in its domain, and adding certain features according to particular observations. The resulting description is stated in terms of a limited number of descriptors. A critical point is that the frame, as the unit of represented knowledge, is quite large. Rather than being on the order of a single property or relation attributed to an object, it is on the order of a description of the object with additional information indicating relations with other frames. Minsky's paper has evoked a great deal of discussion and interest in exploring further levels of detail. It presents plausible and provocative examples of the application of frames to different problems in artificial intelligence. Since then, Winograd (Chapter 7), Bobrow & Norman (Chapter 5), Fahlman (1973), Rubin

(1975), and others have begun to distinguish the various theoretical and technical issues often grouped together in discussions of frames.

In discussions of frames, there is a tendency for supporters of the idea to have an intuitively satisfying internal model of the theory which they have great difficulty making precise and communicating to others. It can be difficult to distinguish clearly between the concept of frames and previous ideas, or even to state the concepts precisely enough to evaluate them at all. In this chapter, I attempt to provide an intuitive model which can serve as a foundation for more precise statements. With an intuitive example in mind, I extract some of the properties which are desirable in a frame representation. Next, I attempt to distinguish those issues which are relevant to recognition, that is, the problem of selecting one of a fixed set of alternate interpretations for a collection of observations. I present an example of recognition (in a tiny world) for which actual technical decisions are made. In the last section, I discuss the simplified model of frame-based recognition used by the example, and outline the limits of its applicability.

II. IMPORTANT PROPERTIES OF FRAMES

Some of the important properties of frames as a representation for knowledge are listed below, to be discussed in more detail later.

Description. A frame provides an elaborate structure for creating and maintaining a description. A primitive element of this description may be expanded to a frame when its internal description becomes of interest.

Instantiation. This is the process by which the frame produces a description of the object being examined by substituting observed for predicted values. Features whose real properties have not been observed are represented by default values.

Prediction. The frame's predicted description can be used to guide the collection of observations for instantiation. It also produces the defaults which substitute for unobserved features.

Justification. Different features of the frame description have different amounts of confidence. Some are clear observations, others are choices among few alternatives, and others are default assignments.

Variation. The dimensions and ranges of possible variation of each feature are limited and specified.

Correction. Anomalies may indicate that the current frame is not correct, and that a different point of view is called for. The frame can analyze the anomaly to select a more appropriate replacement.

Perturbation. For small changes in the observer or the observed, perturbation procedures correct the description without complete recomputation.

Transformation. In case of more significant changes, transformation procedures propose frames suitable for the new situation.

A. Scenario

Consider for a moment an intuitive description of how a frame system might work in the everyday vision process. As you are walking through an unfamiliar house, you come to a normal interior-type door, open it, and walk through. At the moment that you open the door, your (entirely reasonable) expectations have already brought a "room" frame to mind. There is no delay in comprehending the fact that you see four walls, floor, and ceiling, since you already "knew" that they would be there, even without having seen them. Indeed, if these expectations had not been fulfilled, and you had been presented with, say, a seashore instead, you would experience a sense of disorientation. You have found a room, however, and your

(mostly unconscious) analysis continues. The window on the opposite wall is incorporated into the room description which is forming in your mind, very quickly because you have available a number of prepackaged window descriptions. These descriptions are also frames in their own right, but will only be used as stereotypes unless you direct your attention to them. A bed in the room causes the general "room" frame to be replaced by a more specific "bedroom" frame, in which a dishwasher is no longer a serious possibility. The visual information already collected by the "room" frame, however, is still valid and is incorporated into the description within the bedroom frame.

Your attention passes over a clock near the bed and focuses on the fireplace. The fact of its existence and the superficial properties of fireplaces are recorded in the top-level room frame, but another frame is activated to record the description of the fireplace in detail. That information is extraneous in the room frame, and needs a context of its own. When questioned later, you will be able to answer detailed questions about the fireplace (perhaps noticing a subjective feeling of focussing attention on the fireplace and away from the rest of the room when answering), and you will be unable to say more about the clock than that it was a clock mounted on the wall. Quite possibly you will recall it as having hands in spite of the fact that, being a very modern clock, it had none.

In constructing the description of the room, you would have verified in passing that it was a clock, perhaps by noticing the characteristic hour marks, and then allowed the stereotype description of the clock feature to provide the rest. This kind of self-deception by expectation is a result of the diligence of the frame mechanism attempting to extract a maximally detailed description from a minimal amount of input information. I use an example where the default assignment was incorrect because there is less doubt in such cases that the information was supplied by the frame. In general, of course, such stereotypes are correct, making it uncertain whether the information came from a default description or an actual observation.

B. Description

A frame has a small domain of expertise and contains the knowledge necessary to create a description of an object in that domain. Some knowledge tells how to take a set of observations and create a correspondence between those observations and the descriptive mechanism of the frame. Other knowledge allows the frame to predict some features of the description after observing others. Transformation knowledge maintains the description under small changes of viewpoint, to avoid having to redescribe the scene. We can begin to make a distinction between the knowledge in a frame which is about the object being described (the expected features and the relations holding among them), and that which manipulates the description in response to new observations or changes in viewpoint. The latter kind deals with the relations among descriptions and so could be considered as describing the properties of the domain and not of an individual object.

There is an important point to be made about the relation between the local nature of observations and the global nature of descriptions. The global order imposed on the sensory inputs must be learned: it is not intrinsically present in what is seen. Any theory of representation of knowledge, and of recognition in particular, is trying to explain exactly how we impose the order we have learned through experience onto the extremely varied and disordered sensory inputs we receive. The important point, then, is that any global knowledge contained by a description must have come from the internal representation. It could not come from the observations alone. This helps to explain how prior knowledge is not only helpful, but necessary for understanding and perception.

The description of an object includes a number of features of that object (which Winograd in Chapter 7 calls IMPs, for IMPortant elements) and the relations which hold among those features. The description also specifies a limited set from which those features and relations are chosen. It is reasonable to ask about the stove when thinking of a kitchen, but in an "office" frame the stove is not mentioned, not even to say that there is none. The

description may also contain information computed from observations, but which is certainly not in the sensory image--for example, how many people can be served at the dining room table.

C. Instantiation

Instantiation is the process by which a frame creates a description from observations of an object in its domain. Part of a frame is a description schema which makes building a description a matter of making a number of simple decisions and choosing from among limited sets of alternatives. Most of the choices involved in constructing a description have already been made by selecting that frame. For example, we know that virtually all rooms are bounded by plane polygonal surfaces, and that almost all of those consist of six rectangles: four walls, floor, and ceiling. Thus the part of the descriptive process that describes the walls can use a quick, simple test for large deviations from the expected four-wall description. If no deviations are noticed, the complex description of four perpendicular rectangular walls can be used in the particular room description. This process, based on our experience with typical rooms and the appearance of typical room-edges from the usual perspective, makes it possible to verify a complex portion of a description in much less time than would be required to generate it from scratch.

Our actual experience with rooms comes mostly from particular kinds of rooms: rooms in homes, offices, schools, and other buildings. As we instantiate the general "room" frame we record characteristics which could belong to any kind of room. At the same time, however, the features we see specify which particular kind of room is before us, and bring in the frame corresponding to that kind of room. This is the process of refinement: within a frame of common characteristics, making decisions which determine a particular and more specialized frame in which to continue the description. For example, in the scenario, upon noticing the bed, the room frame becomes a bedroom frame, which affects some (not all) of the remaining alternatives in the description.

D. Correction

In most common cases of recognition the identity of the object being described is not initially known, so selecting the proper frame to instantiate is part of the problem. The current "best guess" frame attempts to create a correspondence between what it expects to see and the observations actually available. If it runs into an observation which is incompatible with its domain, that observation can often indicate a good replacement frame. For example, in attempting to recognize a large, four-legged animal, a reasonable guess would be that it is a horse. Small horns, however, are incompatible with a "horse" hypothesis, but strongly suggest a cow. A single large horn would suggest a unicorn. Notice, however, that much of the previously gathered information, such as color and location of various body parts, is valid in any of the three potential frames, and need not be observed anew within the new frame. Fahlman (1973) is carrying on research along these lines, and I discuss these issues in more detail in Section III.E.

E. Default Values

When some feature of a description has not been observed, either because it is hidden or because it simply has not yet been attended to, the frame can still make quite an accurate prediction about that feature. This is true even if the object has not been observed at all yet, and the only basis for prediction is personal, idiosyncratic experience. For example, if I mention a beachball, I immediately conjure up an image of a particular ball with red and white stripes. These default values are very weakly bound features of the description in my "beachball" frame. It would take very little sensory evidence to make me replace them with better data for a particular description. On the other hand, if I see a line drawing of a cube, I have a very strong expectation of a hidden corner and three more faces, and these expectations would be quite hard to replace.

These default values have two quite important uses.

The first is in guiding the process of recognizing and instantiating a particular description by suggesting what features to look for and where to expect them. The second is to provide answers to questions for which observations have not yet been made. In this way, the frame represents our inductive knowledge of the world as gained by previous experience with that domain of objects. This use of default values also allows a frame representation to satisfy the "principle of continually available output" (Norman & Bobrow, 1975), which says that a process should be able to provide a result even when its analysis has not yet been completed. A lack of data or processing resources should produce a graceful degradation of the quality of the output, but not prevent results from being produced at all.

F. Variation

A frame represents a certain limited domain, and hence a range of variation for objects which belong to that domain. As we saw in the room scenario, the features of a frame may be frames in their own right, embodying ranges of variation. On entering a room, you are prepared for certain typical pieces of furniture. A park bench or a diamond-encrusted throne would be outside the permissible range of variation in this frame. Such an anomaly may indicate to the correction mechanism that another frame is called for. When a number of features are near the extremes of their ranges of variation, their collective unlikeliness can cast doubt on the applicability of this frame and initiate a search for further evidence which may result in a new frame being selected. This is particularly clear in medical diagnosis, where a set of symptoms may be possible within the frame for disease X, but so unlikely that the doctor orders further tests to search for a more plausible hypothesis.

G. Perturbation and Prediction

There are a number of different circumstances when a frame may be transformed or replaced by a different one.

While sitting in a room, if I turn my head, I bring a previously invisible region into my field of vision and lose a region from the other side, or I may move, changing the vantage point from which I view certain features. These are relatively small changes which cause perturbations of the frame and the description it produces. I may experience larger changes by walking into another room, requiring a prediction of what frame I may need next and repeating the instantiation process. These phenomena are not isolated, but lie on a spectrum which includes looking from outside the doorway, or lying on the floor and looking up. These intermediate cases include more common information from the original frame than leaving the room entirely, and involve a more drastic change to the frame than a perturbation.

The common element to extract from these transformations is the idea of partially changing a description while saving those portions which are still valid for the new version. A transformation in viewpoint does not take place spontaneously. It occurs as the result of some action (perhaps mental) with which we may be quite familiar; familiar enough, in fact, to be represented by a scenario frame. An action, like an object, has a description, which often takes the form of a scenario. Frames may certainly be used to represent the kinds of variability scenarios are subject to, just as vision frames represent variability in visual descriptions. Part of the frame for a given action will be a prediction of its effect on commonly associated objects and environments. When I am walking, the "walking" frame will predict the change in the visual geometry of the enclosing room. Conversely, strong visual cues can be used very effectively, in movies for example, to evoke the sensation of motion.

When the action in question forces most of the description to be redone, as when I walk from room to room, then the transformation consists mostly of proposing possible new frames. In a familiar house I may be able to summon up a fairly complete and accurate description from memory, but in an unfamiliar house I need time to get my bearings. For small perturbations, however, such as moving slightly within the same room, the visual geometry of the outlines of the room may change slightly, but most

of the features will remain the same, and appear in corresponding places on the walls. If a piece of furniture looks substantially different from the new angle, its own frame may require a transformation.

Occlusion of objects in the background by those in the foreground can be explained by their relative positions within the room description. I do not believe, however, that people accurately predict such occlusions from their mental descriptions. On looking at a scene, the description I generate is not of the picture I see, but of what I think that scene actually is. I come to conclusions about the global nature of the scene from evidence I have, and fill in with default values where I missed actual perceptions.

H. Extreme Anomalies

An extremely unexpected observation, such as opening that door and finding myself at the seashore, is treated in a more serious way. My dumbfoundedness resulting from this occurrence is not only due to the time it takes to find a "seashore" frame, but I am also faced with evidence suggesting that previously accurate notions of continuity no longer hold. I do have some knowledge of geography, and I am filled with curiosity about how I was transported to the sea without noticing. I may decide to reject the evidence and the attack on continuity by concluding that I am dreaming or have gone crazy. Alternatively, I may retreat back through the door and lock it, or in the best Kuhnian tradition, postpone dealing with such questions while I explore and gather more observations. The point of all this is that an extremely unexpected occurence calls into question not only the predicted frames that have proved to be inaccurate, but also that knowledge which led the prediction process so seriously astray. Such experiences are saved and incorporated into newer versions of the faulty frames when structural revisions become possible.

III. AN EXAMPLE: BLOCKS WORLD RECOGNITION

Frames, then, have an intuitive appeal as a metaphor to explain how people organize and represent their knowledge. An obvious question is, of course, whether this idea is of any help to us in representing such knowledge in computer programs. The next example solves a very easy problem, one for which the machinery developed is quite superfluous. The hope, however, is that the way such problems are solved will provide valuable techniques to be used in solving larger, more realistic problems.

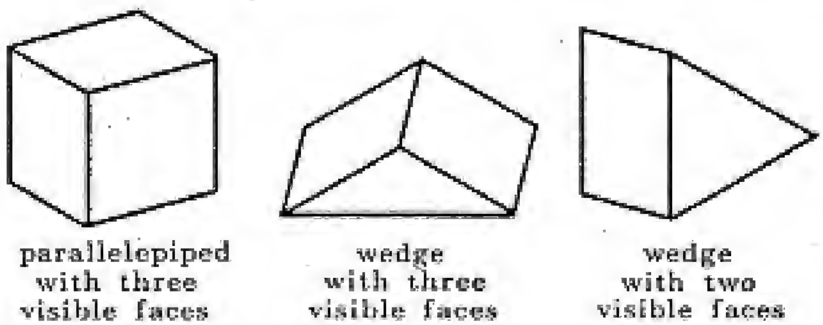

Fig. 1. The domain.

The domain shown in Fig. 1 consists of line drawings of a single, unoccluded block, which can be either a parallelepiped with three visible faces, a wedge with three visible faces, or a wedge with two faces visible. The blocks world has been used as a domain by a number of researchers in different contexts (Winston, 1975; Winograd, 1972), and is rightly criticized as a "toy" world, lacking many of the important and complex problems found in the real world. Much of the difficulty of real world domains comes from our inability to express in a computer program descriptions and distinctions which are obvious (though hard to verbalize) to a human being. The blocks world, however, has very clear descriptive mechanisms, and it is easy to find precise distinctions between two line drawings. In this domain we can focus on the nature of the recognition process, and how the use of frames in

manipulating and representing descriptions can aid that process. The hope is that a simple "toy" example will clarify phenomena which would be obscured by other important (but separate) issues in a more complex domain.

Five of the phenomena mentioned in the previous section will be addressed by the example of the block recognition program: description, instantiation, prediction, correction, and transformation. The recognizer instantiates a description of the object it recognizes, using its predictions to guide the recognition. When a conflict occurs between prediction and data, a complaint department associated with the frame selects an appropriate course of action, often a transformation to a new frame.

What, then, does the recognizer take as its input and produce as its output? The "sensory" world of this system consists of a body of data about the line drawing which can be interrogated by asking it questions which are very local, in the sense that a particular part of the visual scene can be reached only by searching along a known edge from a vertex which has already been observed. An attractive metaphor is that of walking over a snowy field attempting to interpret a line drawing laid in pipes hidden under the snow. More precisely, the sensory world consists of edges and vertices, which can perform the following operations upon receipt of the appropriate message.

> A vertex will deliver its type, the edges which terminate at it, and the sizes of the angles between pairs of edges. This corresponds to the result of a "circular search" in the neighborhood of a vertex. The type of a vertex is L, fork, or arrow. The size of an angle can be described as either acute, right, or obtuse.
>
> An edge will deliver its "other vertex" upon being presented with one vertex. This corresponds to scanning an edge from one vertex to the other.

With this limited sensory world, and even more impoverished descriptive system, the recognizer will attempt to recognize what it sees and provide a global description

of that object. It is important to recognize the difference between the sensory world which is available, and the descriptive mechanism which creates an internal representation to be remembered. Even if the sensory world provided precise angle measurements, the recognizer could only describe them as acute, right, or obtuse. Similarly, people discard or blur many distinctions which are physiologically available to their senses.

What is the description of a line drawing? A description imposes a level of organization on the observational data which is not locally apparent in the scene itself. Simply by stating that an object is, say, a cube, the description asserts that a certain collection of features appears in the scene and that many others do not, a fact which could be determined directly only by exhaustively searching the scene. The description also provides a global structure to the features which is not apparent in the local relations of the scene. Thus, looking at one corner of a cube, one may ask of the description, "Where is the opposite corner?" The scene cannot answer such a question, for it cannot define "opposite" in a way that is meaningful to the cube. A third function of the description is to include properties of the object which are inferred from the observed features along with the knowledge of its identity, such as the volume of a cube or wedge. The description produced by this recognizer will fill only the first two functions, noting collections of features and providing a global relational structure. A line drawing will be classified according to type, and its parts will be accessible according to the global structure of the object it represents.

The recognition problem in this blocks world domain is to select and instantiate the correct frame for the drawing. Since, however, instantiation must begin before selection can take place, the recognizer must also evaluate observed evidence, predict subsequent observations, select a new frame when necessary, and save previously collected observations.

Having defined the problem, we can now begin to look at what the recognizer is. The recognizer consists of three frames, one for each type of object in the domain. Each frame is a program for examining the input data and

constructing a description of its type of block from that data. A frame has many of the properties of a description, in that it imposes its own global organization on the observed data and makes predictions based on its observations along with its assumptions about the type of object being observed. An important similarity between a frame and the corresponding description is that a frame will be able to answer questions about as yet unobserved portions of the scene based on its predictions. Thus a frame functions as a complete (though possibly erroneous) description even before its processing is complete.

A frame, however, has additional capabilities which are not present (or necessary) in a description. It contains strategy knowledge which can advise it on the best observations to consider as it builds its description. It also has the ability to evaluate the observations for consistency with the description it is attempting to instantiate, and to turn the process over to a more appropriate frame when a fatal inconsistency appears. During the recognition process this description serves as a hypothesis about the scene which the frame is attempting to confirm or refute. When the hypothesis is refuted, however, it is not only the description which is replaced by a better alternative. The new frame also contains new knowledge about strategy, evaluation, and the handling of inconsistencies in ways that are more appropriate to the new hypothesis.

There are two distinct kinds of knowledge about the features of these line drawings which are embedded in the frame and which guide the construction of the description. The first is local knowledge about the types of vertices which appear in the figure, and how each vertex is connected to its immediate neighbor. The second is knowledge of the global relations which hold among the angles in different parts of the drawing (see Fig. 3). These global relations allow an observed angle measurement in one part of the drawing to predict an observation in another part. Both kinds of knowledge serve the same role of predicting observations and guiding the recognition process, but they interact with observations in different ways, and the details of their representation in the frame are somewhat different.

A. Recognition Scenario

Let us follow a scenario of the recognition of a block drawing, in this case the three-face view of a wedge. Figure 2 shows the stages of the recognition process, with observed data indicated in solid lines and hypothetical knowledge in dotted lines. The first drawing is the actual scene, with the vertices numbered in the order in which they will be explored.

Vertex 1: We start the recognition process by giving the program an initial vertex, which in this case happens to be an L-vertex. The initial hypothesis is that the figure is a parallelepiped, indicated by the dotted lines in the figure. The single angle measurement, along with the parallelepiped hypothesis, predicts the sizes of the four additional angles indicated.

Vertex 2: The second vertex observed agrees completely with the hypothesis, which expected an arrow vertex and had a particular measurement anticipated for the left side angle of the arrow. The two other angle measurements provided by the arrow allow the frame to predict every angle expected in the parallelepiped. Figure 3 shows the global angle relations which support this extensive prediction.

Vertex 3: This is an arrow vertex, which is the vertex-type predicted by the current hypothesis. At this point we can see that the angle is too small, and that the figure cannot be a parallelepiped. If the program had been given better angle resolution, the angle specialist would also have noticed the error in angle and would have complained to the frame. We are assuming, however that the system cannot discriminate well enough, so the angle specialist accepts the information as consistent, and the recognizer continues with a mistaken hypothesis.

Vertex 4: The fork-vertex at the center of the figure also corresponds completely with the parallelepiped hypothesis. One complete face has now been observed and confirmed.

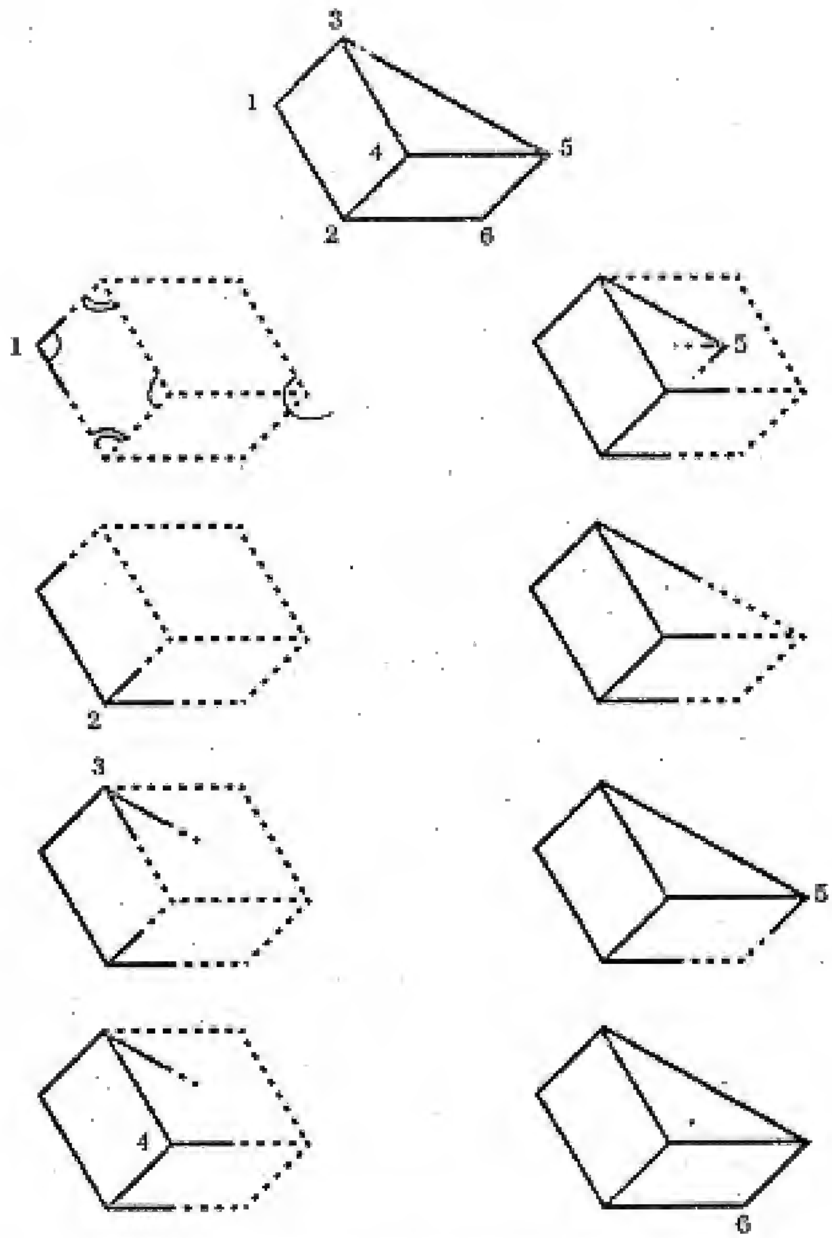

Fig. 2. Stages of the recognition scenario.

Vertex 5: With this observation, the parallelepiped hypothesis finally breaks down. The L-vertex specialist observes an unexpected type of vertex and complains to the frame: "I expected an L, but got an arrow." The parallelepiped frame knows that this particular problem indicates a transition to the three-face view of the wedge. It then analyzes the complaining vertex and the data already collected to discover the correspondence between the cube and wedge frames which will allow previously collected data to be retained. Finally, it executes the selected transformation.

Notice some fancy stepping here. The unexpected arrow vertex was an anomaly to the parallelepiped frame, and the information contained in it could not be completely processed by the L-vertex specialist. Thus it was ignored, and the transition to the wedge frame took place with only the previously known data. Once the new frame was in control, it could deal with the arrow vertex. The arrow vertex, in effect, caused the recognizer to do a "double take".

Vertex 6: At this point, with the three-face wedge frame directing the exploration, there is only one remaining vertex, and it completely confirms this hypothesis. The frame is now fully instantiated.

B. Representation

A frame is built around a hypothetical description. The elements of that description are represented by active program objects (called "specialists") which interact by sending messages to each other. Each vertex in the drawing is represented by a specialist in one of the vertex types: L, fork, or arrow. The properties of that type of vertex are represented by the particular behavior of that specialist. A vertex specialist has pointers to each of the edges terminating at it. An edge is also represented by a specialist with pointers to its two vertices. This network of specialists connected with pointers represents the topological connectivity of the line drawing. The network makes implicit predictions by stating that if a vertex specialist is satisfied with the real (observed) vertex

corresponding to it, then a scan along one of the edges should encounter another real vertex which will satisfy that corresponding vertex specialist. Once an initial correspondence has been established between observation and hypothesis, this constitutes a prediction of all the vertex types and their connections throughout the figure. This prediction is embedded in the structure of the frame, and cannot be changed by incoming data, except by refuting the hypothesis and replacing the frame with another one. New angle predictions, on the other hand, can be freely sent among angle specialists throughout the figure. The relations among the angles in the line drawing (Fig. 3) are represented by angle and relation specialists, who communicate predictions and observations among themselves. By this communication, an angle observation in one part of the figure can affect the prediction in a remote part of the figure. The edges, the faces, and the block as a whole are also represented by specialists, sending messages to each other, whose behavior directs the recognition and instantiation process.

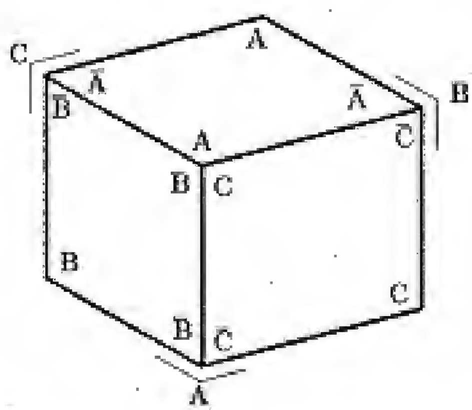

Fig. 3. Global angle relations in the parallelepiped frame.

C. The Basic Loop

The basic operation of the recognition process is to select an observation and evaluate it with respect to the predictions made by the current frame hypothesis. The flow of control described here includes the decisions about search strategy, sending observations to corresponding specialists for evaluation, and communicating predictions and additional data between specialists. It is important to notice that these design decisions can be changed independently. For example, the selection of the next observation can be made in a different way without changing the rest of the flow of control. The range of flexibility of these design decisions will be the topic of the last section of this chapter. Since the frame consists of a number of specialists, each with its own behavior, the description of the normal flow of control will also describe much of the behavior of those specialists.

(1) When instantiation begins, an initial observed vertex is sent to the recognizer. Since the initial "cube" hypothesis is symmetrical, the correspondence between hypothesis and data is set up by sending the observed vertex to an arbitrary vertex specialist of the same type. After this, the specialist for the entire block directs the instantiation.

(2) When the block specialist is told to select an observation, it cycles through its faces, telling each in turn to select the observation.

(3) When a face specialist is told to select an observation, it cycles through its edges, telling each in turn to select an observation. If they all refuse, the face passes the refusal back to the block specialist.

(4) When an edge specialist is told to select an observation, it checks to see if it is in a very particular state. It can make an observation only if: a real edge has been observed corresponding to it *and* exactly one of the vertex specialists at its ends has observed a corresponding real vertex. If this state of affairs obtains, the edge specialist performs the scan from one end of the real edge to the other, and sends the newly observed real vertex to its corresponding vertex specialist; otherwise, a refusal goes back to the face specialist.

(5) When a vertex specialist receives an observed vertex, it evaluates the observation against its prediction, by checking to see if the observed type is the same as what it expected. If not, a complaint goes to the complaint department (more on this in the next section). If the type is acceptable, the vertex specialist obtains the real edges and angle measurements which are available from the observed vertex. It sends the observed edges to the corresponding edge specialists, and the observed angle measurements to the angle specialists.

(6) When an edge specialist receives an observed edge from one of its neighboring vertex specialists, it remembers the real edge, and the real vertex at one end, so it can respond differently to future requests for observations.

(7) When an angle specialist receives an observed angle measurement, it compares the measurement against any prediction it might have. A conflict, of course, results in a complaint sent to the complaint department. If there was no previous prediction, the measurement will be of interest to the specialist (called a "relation") which represents the relation among some collection of angles in the figure, so the observed measurement is sent on. An example of such a relation is that holding between the four angles of a parallelogram.

(8) When a relation receives such measurement, it decides whether this measurement implies some useful prediction. If so, it sends that prediction to the appropriate angle specialists.

(9) When an angle specialist receives such a prediction, it simply remembers it for comparison with future observations.

D. The Complaint Department

A frame has a complaint department which receives complaints about violated expectations from the vertex and angle specialists. The offended specialist sends a description of the problem from its own local point of view, and the complaint department, with its more global knowledge, must select the proper course of action. In this example, only the parallelepiped frame has a nontrivial

complaint department. There are three distinct responses it can make. It can decide that the observed anomaly indicates that the object being recognized is actually the three-face view of the wedge, for example, and that it can determine the correspondence between what has already been observed and the data expected by the three-face wedge frame. The same can happen to indicate a transition to the two-face view of the wedge. The third alternative (Fig. 4) is somewhat more interesting. The complaint department has enough information to decide conclusively that the new frame should be the three-face wedge, but it does not have sufficient data to select the correspondence between the old and the new frames. It cannot decide which face will be the triangle. The solution adopted in this recognizer is to continue the recognition process under the old hypothesis (now known to be mistaken), under the assumption that the next complaint will be able to settle the question. This decision is based on knowledge of the domain which assures the recognizer that no important data will be lost while working under this mistaken hypothesis. I do not address the question of how such knowledge can be automatically acquired from experience.

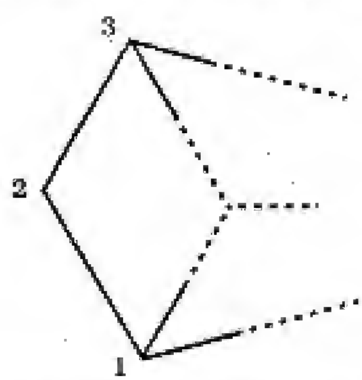

Fig. 4. The ambiguous transition: the frame cannot predict which side will be the triangle.

Table I summarizes the process by which the complaint department deals with anomalies and selects the new frame. The alternative "continue (three-face wedge)" is the case

discussed immediately above. The complexity of the complaint department is a result of the number of complaints which are meaningful, and hence of the number of alternate frames known to this frame. This table does not show the fairly elaborate decision procedure for determining the correspondence between the two frames before the transition can actually be executed.

TABLE 1.
The Complaint Department

```
Vertex specialist
        expected arrow, got L ==> two-face wedge
        expected L, got arrow  ==>  three-face wedge

Angle larger than expected ==> three-face wedge

Angle smaller than expected:
  in L vertex  ==>  two-face wedge
  in arrow vertex:
    full angle  ==>  continue (three-face wedge)
    side angle:
      observed L vertex in that face  ==>  two-face wedge
      else  ==>  continue (three-face wedge)
  in fork vertex  ==>  three-face wedge
```

E. The Transition

Once an anomaly has refuted the parallelepiped hypothesis, and a more appropriate wedge frame has been selected, the problem remains of actually performing the transition. The simplest solution would be to start over, ignoring previously collected data except to indicate a different frame to start with. This form of recognition is a blind, back-tracking search through a space of line drawings. One goal of this example, however, is to show

how frame-based recognition can exploit the similarities between different line drawings to preserve observations collected under a mistaken hypothesis. At the very least, the actual observations of edges and vertices can be mapped from the old description to the new one because the definitions of adjacency and connectivity are shared by all frames in this domain. In favorable circumstances, higher-level descriptive objects, such as a parallelogram face, will remain valid in the new frame without disturbing their internal structure.

The transition from the parallelepiped frame to the three-face view of the wedge has these favorable properties. The differences between the two descriptions are confined to changing one parallelogram face to a triangular face, and adjusting the angle predictions. To accomplish this transition, the parallelepiped frame replaces the collection of specialists which represent one parallelogram face with another collection for a triangle. It transfers whatever data has already been observed to the corresponding new specialists, notifies all concerned neighbors of the change, and the displaced parts of the old description disappear. The internal structure of the neighboring faces changes only in accepting a new pointer. The angle predictions also vanish, but new predictions are solicited from the angle specialists.

The transition to the two-face wedge is quite different, however. The change here involves much more extensive changes to the structure of the description. Just as in the other transition, there is a correspondence between the representing specialists in the two frames, but in this case specialists who correspond may not have the same behavior. Faces which had two neighbors now have only one; vertices which expected to be arrows will now be Ls; and as before, the angle predictions become obsolete. In this case, all that can be salvaged from the old frame are the actual observations, including the connections between them. These observations are transferred to corresponding specialists in the two-face wedge frame, which incorporates its own higher-level descriptive structure. There is still an important saving in observations to be investigated, but not as much program structure can be shared between the two-face wedge and the parallelepiped as was possible between the parallelepiped and the three-face wedge.

F. The Implementation

This example was first programmed and hand-simulated in ACTORS (Smith & Hewitt, 1974). As the ideas continued to evolve, a working implementation in SMALLTALK (Kay, 1974) was written and debugged in less than two weeks. The ease with which the concept could be translated into a working program is primarily due to the novel semantics of these two languages. Both ACTORS and SMALLTALK evolved from the ideas in SIMULA (Dahl & Hoare, 1972), and are what might be called actor languages, as opposed to function or procedure languages like LISP or ALGOL. An actor is a procedure which can maintain an internal state between invocations. Actors communicate by sending messages to each other, and are not constrained to send messages (or control) only up or down a function-call hierarchy. Allowing an actor to maintain an internal state makes it possible for the variables which are intuitively associated with a conceptual object to be associated directly with the corresponding program object.

A certain amount of confusion is possible between the different types of instantiation in this example. A specialist representing a feature of the line drawing (for example, an arrow-vertex specialist) is written as an actor which maintains a certain amount of internal state, and has a certain behavior in response to particular messages. The parallelepiped frame contains three copies of the arrow-vertex specialist, each of which is an instance of the actor mentioned above. These three instances are not identical, but can be distinguished by which other specialists they have as neighbors. The parallelepiped frame, then, is a program which consists of several parts, some of which share program text but have different internal states. This frame is then provided with a source of observational data. Instantiation of the description is the process by which the various parts of the frame establish a correspondence with observational data. To add further to the confusion, we can imagine a scene containing two unoccluded blocks, for which we make two copies (instances) of the entire recognizer, so that separate frames can be instantiated, resulting in two independent descriptions. This third case seems to have no theoretical interest.

IV. WHAT DOES THIS ALL MEAN?

Let us step back now and see what significance this example has in the larger enterprise of representing knowledge for recognition. The overall structure of the recognizer has some applicability to other domains in which greater expressive power is required of the descriptive mechanism. In the following sections, I discuss the general conclusions which can be drawn about the descriptive mechanisms used, and about the interacting modules which supervise the recognition and instantiation processes. Other domains which have been investigated in some depth, and from which I draw examples, are medical diagnosis (Rubin, 1975), and electronic circuits (Sussman, 1973). These other domains can show features which fit into the framework I have developed, but which do not appear in the blocks world. Where possible, I point out the range of applicability of this framework for recognition, and give examples where it does not apply.

A. Representing the Hypothesis

The block recognizer uses three methods to represent hypotheses about line drawings. They are:

* the vertex-specialists, which know about a particular type of vertex to which they expect to correspond;

* the network of neighbor pointers, which links the edge- and vertex-specialists, and homomorphically represents the connectivity of the edges and vertices in the drawing;

* the angle specialists, which represent the global relations among the angle measurements, and actively communicate predictions about particular angles.

This division of representational effort works in the blocks domain because a clear distinction can be made between the different properties to be represented. There are strictly local features (the vertex types), fixed global relations (the connectivity between vertices and edges), and predictive global relations (the angle relations).

Certain other domains fit into the same descriptive framework so that this distinction between local and global features can be clearly made. A good example of this is the domain of electronic circuits, where the connectivity and local properties of components must also be represented, and global relationships among current and voltage measurements at different points can be predicted. A less geometric example with the same logical structure might be representing the time course of certain diseases, where local specialists are able to recognize particular symptoms, the network of connections is a partial time ordering of events, and the global relations may be among the different measurements of a varying quantity, such as blood pressure or white blood cell count.

There are, of course, many domains where the representational structure described in this example does not clearly apply. This is particularly true when features are not as discretely separable as they are in the blocks world. For example, in medicine it can be important to describe the onset of a certain symptom as "insidious", or otherwise specify an indefinite time interval which can overlap with other events. Notice that we are not simply specifying an interval whose endpoints are discrete (though currently unknown), but rather an interval which fails to possess definite endpoints. The network representation described above lacks the expressive power to deal adequately with this phenomenon.

B. Manipulating the Hypotheses

In the previous section, we saw what kind of expressive power is available for representing hypotheses to this kind of recognizer. Now let us consider the structure provided to manipulate those hypotheses. It consists of four parts:

* a module to select the next observation to consider;

* a module to evaluate the observation, comparing it with what was predicted;

* a module to serve as a complaint department, deciding what to do in response to an observed anomaly;

* a module to perform the transition to a new frame, preserving as much as possible of the old information.

These "modules" do not correspond to segregated pieces of program in the block recognizer, but are design units whose implementation is likely to be distributed among the specialists which comprise the frame.

In the following sections, we will examine these modules individually and see what range of behavior can be expected of them. The important questions to ask of each one are: What is it asked to do? What knowledge can it consider? What answers can it give? This modularized view of the recognition process also has its limitations, again because of the discrete structure of frames linked by explicit transitions. This simplified view of recognition is based on the assumption that recognition proceeds by adopting a single "best guess" hypothesis, and modifying it to a better one in response to an unexpected observation. There is no provision for entertaining several different hypotheses at once, or for leaping to an unrelated frame where no explicit link exists. There are also important questions about sharing knowledge among distinct frames which are not addressed in this domain.

C. Selecting the Next Observation

This module decides which potential observation would be most useful at each point in the recognition process. Once it has selected one, it sends the observed data to the appropriate specialist to begin the evaluation. The interesting thing about this module is the range of information it can consider, and where it obtains that information. In the block recognition example a particular set of considerations is designed into the selector, so it does not answer the questions below each time it makes an observation. Doctors, on the other hand, are trained to ask these questions explicitly in the course of a medical examination.

* Given what has already been observed, which alternate hypotheses are the most likely? (i.e. for differential diagnosis)

* The frame uses observed data to refine its predictions and the description it is producing. Which observations would be most productive at this time?

* The pragmatic context of this recognition act makes certain parts of the description more useful than others. Which are these?

* What costs (e.g. pain, risk, money, doctor's time) are associated with potential observations?

There are, of course, some cases where the relative importance of these factors may be decided once and for all and designed into the selection procedure, and others where the situation must be actively and frequently reevaluated. Differential diagnosis information may be requested by the complaint department in cases where an anomaly has been observed, but a unique replacement hypothesis cannot be selected.

D. Evaluating the Observation

This evaluation is a point of close contact between the representation and the manipulation of the hypothesis. The frame checks an observation against its hypothesis, asking whether that observation is consistent with the predicted description. The discussion of representation above illustrates the local nature of this evaluation in the block recognizer. The appropriate vertex and angle specialists each check the consistency of the new information with their expectations. The complexity lies in the range of potential results of this evaluation. In the block recognizer, only the first two of the following possible reactions can occur.

* The observation is consistent with the hypothesis, perhaps providing additional information to be absorbed by the frame.

* It is inconsistent, refuting the hypothesis, and the specialist sends a description of the problem to the complaint department.

* It is consistent with the current hypothesis, but singles out a special case about which more knowledge is available.

* It, in isolation, is consistent with the current hypothesis, though near the edge of the range of variation. However, enough other observations are also near the edges of their ranges of variation that the frame becomes suspicious and complains to the complaint department.

The third, or "further specification" link between frames provides additional information which allows more detailed predictions or better selection of observations. The fourth possibility allows suspicion to be cast on a hypothesis as it becomes more and more unlikely, even though it may never be conclusively refuted. It may nonetheless be replaced by a better alternative.

E. Selecting a New Hypothesis

This module, the complaint department, is given a description of the current complaint (and perhaps remembers past ones), and is asked to select a new hypothesis. In the block recognizer, most of the possible anomalies simply specify unambiguously the frame which should replace the current one. As we saw above, however, there are cases in which further information is necessary to select the correct orientation for the new frame. In either mode the complaint department must possess knowledge about which alternate hypotheses are available. In most cases of practical recognition these decisions will be reduced to simple tests of the observations, just as in the block recognizer, rather than active problem-solving during the recognition process. The speed of frame-based recognition depends on the assumption that the number of potential alternatives in a domain is manageable, and that most anomalies clearly suggest alternate hypotheses.

The eventual answer provided by the complaint department should be a new frame to replace the complaining one. Some of the potential courses of action leading to this result are:

* the anomaly may simply specify a new hypothesis to replace the old;

* there may be previously collected information which can be reexamined in more detail to decide between potential new frames;

* the complaint department may request a particular observation for differential diagnosis from the module which selects the observations;

* if the anomaly is minor, or there are no good alternatives, the current frame may just remember the problem and continue under the old hypothesis, hoping that further observation will illuminate the situation.

The complaint department is also involved in representing the frame's range of variation. Each feature of the frame description has its own range of variation which it will accept before complaining. The complaint department may then decide to disregard certain complaints or accept excuses under some circumstances. A frame system could believe that all dogs have tails, but admit the possibility that a dog without a tail could still be a dog.

F. Translating Knowledge to the New Hypothesis

At this early, somewhat speculative stage of research, it is considerably harder to generalize about the transition procedures than it is to talk about the other parts of the recognizer. The other parts of the recognition process depend largely on the properties of the domain; the transition depends on the structure of the description. Since that structure is one of the goals of our research, any conclusions drawn from it are necessarily tentative. Another caveat is that the blocks world domain was deliberately chosen to minimize the complexity of the descriptive and expressive problems to be encountered.

As I mentioned previously, the hierarchical structure of the description is important in determining how much can be saved in replacing one frame with another. When a

large, self-contained substructure such as a parallelogram face is essentially the same in the two descriptions, it is natural to preserve it as a unit rather than reconstructing it in the new frame. Even more than this is true in the transition to the three-face wedge: only a few parts of the top-level description need to be changed. The rest of the description remains the same.

In making the transition to the two-face wedge, the higher structures of the two descriptions are quite different, so less of the old description can be preserved. The interpretation of the observations remains the same, however: if two parts are considered connected by the parallelepiped frame, they are connected in the wedge frame, and the terms in which they are described are the same. Thus when the recognizer realizes it is looking at a wedge, it can remember what it saw when it thought the object was a parallelepiped. Even when the higher-level descriptive structure must be replaced, the recognizer need not look again at features it has already observed.

Here again we see an example where we are helped out by the good behavior of the domain, or at least of our view of the domain. Even when the frame changes, the interpretation of the observations remains much the same. This need not be true in domains with segmentation problems. For example, in speech recognition, changing the interpretation of one segment may affect the boundaries of the segment, requiring changes which ripple outward to neighboring hypotheses. A different set of techniques is required to state and evaluate hypotheses about domains where segmentation is an important problem.

V. SUMMARY

In this chapter we presented the idea of frames in a very intuitive way, outlining a number of desirable features of a representation for knowledge, and illustrating them with a specific example from the blocks world. A frame is a specialist in a small domain. It contains the knowledge necessary to create a description of an element

of its domain from observed data. The features of such a description may be frames in their own right, representing a range of variation permitted in that domain. The frame for an object can have associated with it frames for actions which commonly affect that object, so that predictions can be made about required modifications to the description. The frame is capable of predicting unobserved features, and of using previous observations to refine its predictions. These predictions can guide the recognition process, and provide answers to questions before that process is complete. An observation which is inconsistent with the frame's expectations can suggest a better frame as a replacement. Much of the partially constructed description can be incorporated into the new frame, which continues the recognition process.

It is important to recognize the value of the intuitive model presented above. In a sense it is a "wish list" of desirable properties for a representation, but it is a list compiled with the problems of effective computability in mind. It will be many years before the technical problems implied by a frame theory can be precisely stated and solved. Such intuitions are therefore all the more important for providing a context in which current research can be viewed.

ACKNOWLEDGMENTS

I have received extremely valuable comments and suggestions from Marvin Minsky, Daniel Bobrow, and Carl Hewitt. Other people who have been very helpful include Michael Dunlavy, Mitch Marcus, Robert Metcalfe, Keith Nishihara, Don Norman, and Terry Winograd.

This research was done in part at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-70-A-0362-0003. I am also grateful to the Xerox Corporation for a stimulating summer spent at their Palo Alto Research Center.

REFERENCES

Dahl, O. J., & Hoare, C. A. R. Hierarchical program structures. In O. J. Dahl, E. W. Dijkstra, & C. A. R. Hoare (Eds.), *Structured programming*. New York: Academic Press, 1972.

Fahlman, S. E. A hypothesis-frame system for recognition problems (MIT-AI Working Paper 57). Cambridge: M.I.T., December, 1973.

Kay, A. SMALLTALK: A communications medium for children of all ages. Palo Alto, California: Xerox Palo Alto Research Center, Learning Research Group, 1974.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Norman, D. A., Bobrow, D. G. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, *1*, 44-64.

Rubin, A. D. Hypothesis formation and evaluation in medical diagnosis (MIT-AI Technical Report 316). Cambridge: Massachusetts Institute of Technology, 1975.

Smith, B., & Hewitt, C. Towards a programming apprentice. *Artificial Intelligence and Simulation of Behaviour Summer Conference Proceedings*, University of Sussex, July, 1974.

Sussman, G. J. A scenario of planning and debugging in electronics circuit design (MIT-AI Working Paper 54). Cambridge: M.I.T., December, 1973.

Winograd, T. *Understanding Natural Language*. New York: Academic Press, 1972.

Winston, P. H. (Ed.). *The psychology of computer vision*. New York: McGraw-Hill, 1975.